

平成 30 年 6 月 27 日現在

機関番号：14603

研究種目：研究活動スタート支援

研究期間：2015～2016

課題番号：15H06344

研究課題名(和文)複数言語で開発されたウェブアプリケーションからのコードクローン検出

研究課題名(英文) Detecting Interlanguage Clones for Multilingual Web Applications

研究代表者

崔 恩瀾 (Choi, Eunjong)

奈良先端科学技術大学院大学・情報科学研究科・助教

研究者番号：90755943

交付決定額(研究期間全体)：(直接経費) 2,200,000円

研究成果の概要(和文)：本研究では、複数言語で開発されたウェブアプリケーションからコードクローンを自動的に検出するツールを開発した。開発したツールの有効性を評価するために、複数言語で開発されたオープンソースのウェブアプリケーションに対して実験を行い、以下を明らかにした。

- 1) ツールが検出したコードクローンは正解か
- 2) コードクローンであるのに、本ツールにより検出できないコードクローンは存在する

研究成果の概要(英文)：In this study, a tool that detects interlanguage code clones from multilingual web applications was developed. To evaluate the effectiveness of the developed tool, the experiment was conducted to open source software projects and the accuracy of the tool has been confirmed. Moreover, the tool was improved by investigating causes of incorrect detection and omission detection.

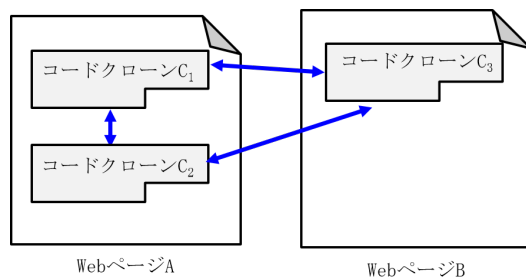
研究分野：ソフトウェア

キーワード：コードクローン

1. 研究開始当初の背景

近年、Web アプリケーションの開発が盛んでいる。Web アプリケーションの代表的な特徴として、1 つの機能を実装するために複数の言語を用いることがある。例えば、Web アプリケーションである機能を実装する際は、クライアント側の実装とサーバ側の実装で異なる言語を用いる。また、Web アプリケーションには、異なる言語で記述されたソースコード間で呼出し関係存在する。従って、開発者がデバッグや機能拡張などの Web アプリケーションの保守作業を行う際には、複数言語で記述されたプログラム、および異なる言語で開発されたプログラム間の呼出し関係を理解する必要がある。しかし、複数言語で記述されたプログラムを理解して、異なる言語で開発されたプログラム間の呼出し関係を把握しておくことは開発者にとって困難な作業である。

コードクローンとは、ソースコード中に存在する互いに一致もしくは類似した部分を持つコード片のことであり、主にソースコードのコピーアンドペーストによって生成される。一般的に、コードクローンはソフトウェアの保守作業を困難させる要因の 1 つである。例えば、以下の図でコードクローン C1 を編集する場合、そのコード片に対応する全てのコードクローン C2、C3 に対しても一貫した編集の是非を検討する必要がある。従って、ソフトウェアでコードクローンを自動的



に検出し、それらのコードクローンの効率的に管理する必要がある。

Web アプリケーションは類似サービスを提供する 17-63%のコードクローンを含んでいる。また、それらのコードクローンを集約することで、Web アプリケーション保守性を向上することが期待できるこれまで提案されてきた多くのコードクローン検出ツールは主に、Java や C などの単一のプログラミング言語で記述されたソフトウェアからコードクローンを検出する。そのため、これらのツールを用いて Web アプリケーションでコードクローンを検出する場合、以下の問題が発生する。

- (1) 複数の言語で記述されたファイルは構文解析できないため、コードクローン検出に失敗する。
- (2) あるプログラムから別の言語で記述さ

れたプログラムを呼び出すことにより、1 つのサービスを実現することがあるが、既存のツールは記述言語毎に検出を行うため、類似したサービスをクローンとして検出できない。

2. 研究の目的

既存のコードクローンの検出ツールを用いて Web アプリケーションからコードクローンを検出する時に生じる 2 つの問題を解決し、Web アプリケーションドから単一および複数言語で記述されたコードクローンを自動的に検出するツールを開発することである。また、開発したツールの有効性の評価を行い、以下を明らかにする。

- 1) ツールが検出したコードクローンは正解か
- 2) コードクローンであるのに、本ツールにより検出できないコードクローンは存在するか、また、なぜ、そのコードクローンは検出できなかったか

3. 研究の方法

Web アプリケーションを対象としたコードクローン検出ツールの実装を行う。本ツールは複数言語で実装された単一ファイルに対して、単一ファイルが単一の言語で実装されるように分割し、ウェブアプリケーションに含まれるプログラム間の静的コールグラフを構築した。また、構築されたコードグラフから同型部分グラフの同値類を検出することによってコードクローンを検出する。

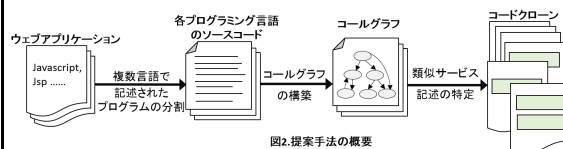


図2.提案手法の概要

(1) 複数言語で記述されたプログラムの分割

Web アプリケーションでは、単一ファイルに複数の言語のプログラムが含まれることが多い。そのため、複数言語で実装される単一ファイルに対して単一ファイルが単一の言語で実装されるように分割を行う。複数言語で記述されたファイルの構文解析の実装は、Island Grammar に基づく方法を採用する。

(2) コールグラフの構築

Web アプリケーションに含まれるプログラム間の静的コールグラフを構築する。各ノードは元々単一言語で記述された、もしくは(1)における分割によって作成されたプログラムファイルの 1 つである。

(3) 類似サービス記述の特定

呼出しグラフから同型部分グラフの同値類を検出することにより類似サービスを特

定する。ノード(プログラムファイル)間において、ホワイトスペースや識別子名の差異を除いて等価であれば、それらノードは等価であるとする。プログラムファイル単位の高速度なクローン検出技術は、ノードの等価性判定については容易に実装することができる。

呼出しグラフの次数が大きくなり、同型部分グラフの同値類の検出が現実的な時間で終了しない場合はまとまりのある成分(弱連結成分など)に分割し、同型グラフの判定を分割後の成分間においてのみ行うようにすることで、計算量の削減を行う。まとまりのある成分への分割については、プログラム依存グラフから同型部分グラフを検出するコールドクローン検出手法を用いる。

4. 研究成果

Web アプリケーションを対象としたコードクローン検出ツールを開発した。また、本ツールの有効性の評価するために、オープンソースのWeb アプリケーションに対して開発されたツールを適用し、本ツールのコードクローン検出精度を確認した。また、ツールの誤検出および検出漏れの原因を調査し、開発したツールの改善を行った。

5. 主な発表論文等

(研究代表者、研究分担者及び連携研究者には下線)

〔雑誌論文〕(計 1 件)

上村 恭平, 森 彰, 藤原 賢二, 崔 恩瀨, 飯田 元: "ハードウェア記述言語におけるコードクローンの定量的調査", 情報処理学会論文誌, Vol.59, No.4, pp.1225-1239, 2018.
<http://id.nii.ac.jp/1001/00187318/>

〔学会発表〕(計 10 件)

Kisub Kim, Dongsun Kim, Tegawendé F. Bissyandé, Eunjong Choi, Li Li, Jacques Klein, Yves Le Traon: "FaCoY - A Code-to-Code Search Engine", Proceedings of the 40th International Conference on Software Engineering, pp.946-957, Gothenburg, Sweden, May-June 2018

徳井 翔梧, 吉田 則裕, 崔 恩瀨, 井上 克郎: "局所性鋭敏型ハッシュを用いたコードクローン検出のためのパラメータ決定手法", 電子情報通信学会技術研究報告, Vol.117, No.477, SS2017-74, pp.57-62, 沖縄産業支援センター, 2018年3月。

Yuichi Semura, Norihiro Yoshida, Eunjong Choi, Katsuro Inoue: "CCFinderSW: Clone Detection Tool with Flexible Multilingual Tokenization",

Proceedings of the 24th Asia-Pacific Software Engineering Conference (APSEC 2017), pp.654-659, Short Paper, Nanjing, Jiangsu, China, December 2017.

沼田 聖也, 吉田 則裕, 崔 恩瀨, 井上 克郎: "開発作業のモニタリングによるコードクローン集約支援環境の構築", 情報処理学会研究報告, Vol.2017-SE-197, No.3, pp.1-6, 熊本大学, 2017年11月。

山田 悠貴, 崔 恩瀨, 吉田 則裕, 飯田 元: "凝集度を用いたメソッドのインライン化の支援手法", 情報処理学会研究報告, Vol.2017-SE-197, No.4, pp.1-6, 熊本大学, 2017年11月。

堤 祥吾, 吉田 則裕, 崔 恩瀨, 井上 克郎: "プログラミングコンテスト参加者を対象とした編集作業の特徴調査", 情報処理学会研究報告, Vol.2017-SE-197, No.6, pp.1-8, 熊本大学, 2017年11月。

石津 卓也, 吉田 則裕, 崔 恩瀨, 井上 克郎: "コードクローンに対するリファクタリング可能性に基づいた削減可能ソースコード量の調査", 情報処理学会研究報告, Vol.2017-SE-197, No.7, pp.1-8, 熊本大学, 2017年11月。

横井 一輝, 崔 恩瀨, 吉田 則裕, 井上 克郎: "情報検索技術に基づくブロッククローン検出", 情報処理学会研究報告, Vol.2017-SE-196, No.19, pp.1-8, 函館コミュニティプラザ, 2017年7月。

瀬村 雄一, 吉田 則裕, 崔 恩瀨, 井上 克郎: "多言語対応のための字句解析機構を持つコードクローン検出ツールの開発", 情報処理学会研究報告, Vol.2017-SE-196, No.20, pp.1-8, 函館コミュニティプラザ, 2017年7月。

Yuta Nakamura, Eunjong Choi, Norihiro Yoshida, Shusuke Haruna, Katsuro Inoue: "Towards Detection and Analysis of Interlanguage Clones for Multilingual Web Applications", in Proceedings of the 10th International Workshop on Software Clones (IWSC 2016), pp.17-18, Osaka, Japan, March 2016.

〔図書〕(計 0 件)

〔産業財産権〕

出願状況(計 0 件)

取得状況（計 0 件）

6 . 研究組織

(1)研究代表者

崔 恩滯 (Eunjong Choi)
奈良先端科学技術大学院大学・情報科学研究科・助教
研究者番号：90755943

(2)研究分担者

(3)連携研究者

(4)研究協力者